ELSEVIER

Journal of Clinical Epidemiology ■ (2014) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Application of credibility ceilings probes the robustness of meta-analyses of biomarkers and cancer risk

Stefania I. Papatheodorou[a], Konstantinos K. Tsilidis[b], Evangelos Evangelou[b], John P.A. Ioannidis[c,d,e,*]

[a]*Department of Health Sciences, Cyprus Institute for Environmental and Public Health, Cyprus University of Technology, Eirinis 95 Str, 3041, Limassol, Cyprus*
[b]*Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, University Campus, 45110 Ioannina, Greece*
[c]*Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine, Stanford 94305, CA, USA*
[d]*Department of Health Research and Policy, Stanford University School of Medicine, Stanford 94305, CA, USA*
[e]*Department of Statistics, Stanford University School of Humanities and Sciences, Stanford 94305, CA, USA*

Accepted 4 September 2014; Published online xxxx

Abstract

Objectives: Meta-analyses of biomarkers often present spurious significant results and large effects. We applied sensitivity analyses with the use of credibility ceilings to assess whether and how the results of meta-analyses of biomarkers and cancer risk would change.

Study Design and Setting: We evaluated 98 meta-analyses, 43 (44%) of which had nominally statistically significant results. We assumed that any single study cannot give more than a maximum certainty $100 - c\%$ (c, credibility ceiling) that the effect estimate [odds ratio (OR)] exceeds 1 (null) or 1.2.

Results: Nominal statistical significance was maintained for 21 (21%) meta-analyses, for $c = 10\%$ and OR >1, and these proportions changed to 7%, 3%, and 6% with ceilings of 20%, 30%, and 40%, respectively. For ceilings for OR >1.2, the respective proportions were 37%, 21%, 7%, and 3%. Seven meta-analyses on infectious agents retained statistical significance even with a high ceiling of $c = 20\%$ for OR >1.00. Meta-analyses without other hints of bias (large between-study heterogeneity, small-study effects, excess significance) were more likely to retain statistical significance than those that had such hints of bias.

Conclusion: Credibility ceilings may be helpful in meta-analyses of biomarkers to understand the robustness of the results to different levels of uncertainty. © 2014 Published by Elsevier Inc.

Keywords: Meta-analyses; Biomarkers; Cancer; Credibility ceiling; Predictive intervals

1. Introduction

Meta-analysis is widely used to synthesize evidence from observational studies [1]. A growing literature of meta-analysis of observational studies on prognostic and other biomarkers suggests that biases may exist that favor the reporting of exaggerated associations with significant results, often, but not exclusively, in small studies [2–7]. Biased results may underlie many of the failures to translate biomarkers into effective clinical tools [8]. Because of the large amount of data synthesized in each meta-analyses, the confidence intervals (CIs) become spuriously narrow and underestimate the biases that may exist but often remain unmeasured in the combined nonrandomized studies. Various methods have been proposed to approach this problem, mainly based on Bayesian statistics [9], assuming that different levels of bias affect the point estimate or the variance [10–14]. Many of these approaches are promising but can be quite convoluted to use and may require meticulous, explicit modeling of each potential source of bias [15]. Modeling each potential source of bias is a very complicated and difficult process. There are only a handful of articles in the biomedical literature where this has been done, and, even then, there is no validation or reassurance that it has led to a more accurate result than analyses that were not corrected for biases.

Conflict of interest: None.

Funding: S.I.P. was funded from a Cyprus University of Technology internal grant (starting grant No 200-099). K.K.T. was supported by the seventh framework program of the European Union for his work on this article (PIEF-GA-2010-276017).

* Corresponding author. 1265 Welch Road, Medical School Office Building Room X306, Stanford, CA 94305, USA. Tel.: (650) 725-5465.

E-mail address: jioannid@stanford.edu (J.P.A. Ioannidis).

http://dx.doi.org/10.1016/j.jclinepi.2014.09.004
0895-4356/© 2014 Published by Elsevier Inc.

What is new?

Key findings

- The application of credibility ceilings in meta-analyses of biomarkers and cancer risk led to many associations losing their statistical significance.

What this adds to what was known?

- Infectious biomarkers were most likely to remain statistically significant even with substantial credibility ceilings being applied.
- Meta-analyses without other hints of bias were more likely to retain statistical significance than those that had such hints of bias.

What is the implication and what should change now?

- Results of meta-analyses of biomarkers should be interpreted with adequate caution and may be downplayed under the assumption of different credibility ceilings for single studies.

We have previously proposed [16] a method for addressing this problem using a simple adjustment of the uncertainty of the results of the included studies without the need to make specific assumptions about the various types of implicated bias. The only assumption is that because of methodological limitations in observational studies, regardless of sample size and quality, a single study cannot offer more than a maximum certainty that an effect is in a particular direction (rather than null or even in the other direction). The maximum certainty is $100 - c\%$, where c is called the credibility ceiling.

The aim of the present article was to extend this method and apply it not only in one meta-analysis but also across many meta-analyses in a specific research field. Moreover, we aimed to examine also credibility ceilings that restrict the maximum probability of the effect being larger than a given nonnull value, rather than larger than the null. This approach may fit well in many observational epidemiology fields, where biomarkers are involved. Many biomarkers may have some small or modest association/prognostic effect, whereas few of them have been rigorously, reproducibly shown to have large association/prognostic effects. We have applied these methods to a previously collected set [5] of 98 meta-analyses of biomarker associations with cancer risk. The changes in the results are then interpreted according to the adjustment for different credibility ceilings. Finally, we also assessed whether meta-analyses that retain their statistical significance despite increasing credibility ceilings are more likely to be those that do not have other hints of bias, for example, small-study effects [17], excess significance [18], or large between-study heterogeneity [19] and whether inferences using credibility ceilings may be commensurate with inferences obtained by the calculation of 95% predictive intervals that accommodate the uncertainty due to the observed heterogeneity across different studies [20].

2. Methods

2.1. Study identification

Details on how the eligible meta-analyses were retrieved have been reported elsewhere [5]. In brief, we systematically searched PubMed from 1966 through December 20, 2010, to identify meta-analyses of observational studies on the association between biomarkers and cancer risk. The search strategy used the keywords "cancer or carcinoma or tumor or neoplasm or neoplasia or maligna*," and they had to appear in the title of the article. When more than one meta-analysis on the same research field was eligible, the meta-analysis with the largest number of studies was retained for the main analysis.

2.2. Credibility ceiling method

If we consider the synthesis of n studies, each with an effect size y_i with variance v_i under a random-effects meta-analysis model, the likelihood of each study will be $L_I(\mu,\theta_i,\tau^2|y_i,v_i)$, a function of the true underlying random effect θ_i, the common mean effect μ, and the heterogeneity variance τ^2. The likelihood function of the meta-analysis model will be $L(\mu,\theta,\tau^2|y,v)=\prod_{i=1}^{n}L_i(\mu,\theta_i,\tau^2|y_i,v_i)$. In the inverse variance meta-analysis framework, $\widehat{\mu}=\sum y_i w_i/\sum w_i$ and $w_i=(v_i+\widehat{\tau}^2)^{-1}$ [21]. The random-effects model inflates the variance of the pooled effect $\widehat{\mu}$ to incorporate heterogeneity, but the CIs may still be insufficiently narrow to include all the factors of uncertainty and bias.

As previously described [16] in the credibility ceiling method in a single study, no matter how large and thoroughly conducted, there is at least a probability c that the effect will not exceed a certain value X. A credibility ceiling may be seen as a cumulative likelihood ceiling under the situation where for each study we do not consider any additional external information about the measure being estimated and simply focus on the data from the study per se. Limitations in the design and the topic under study do not allow c to be reduced further. The value of X can be any number, and in many traditional observational associations of interest with treatment implications, the main question is whether there is any effect of whatever magnitude. For biomarkers, we apply the credibility ceiling method for values $X = 1$ (null effect) and $X = 1.2$ in an odds ratio (OR) scale. The reason for this is that in the case of biomarkers, many biomarkers have a small or modest prognostic effect, but far fewer may have a sizable effect.

For consistency, we used the symmetry of the ORs and reversed the ORs of all studies in all meta-analyses where the summary OR was originally less than 1.00 (decreased risk of cancer); thus, all summary effects were coined to have values ≥ 1.00.

According to the credibility ceiling stipulation, a single study cannot give us more than $(100 - c)\%$ certainty that the effect exceeds each of the plausible values of X. If this level of certainty must be exceeded, several studies replicating an effect greater than X have to be documented. Thus, we consider a random variable $u_i \sim N(y_i, v_i)$ and the probability $P(u < X | y_i > X)$ that the variable takes values less than X is calculated. If this probability is less than c, the variance v_i^* is recalculated as $v_i^* = \max[(y_i/z_c)^2,\ v_i]$ with z being the inverse of the cumulative normal distribution. The new study-specific likelihood functions with the inflated variances are then synthesized according to inverse variance meta-analysis methods, and the variance of the summary effect estimate is further inflated as $\mathrm{var}(\widehat{\mu}) = 1/\sum\left(v_i^* + \widehat{\tau}^2\right)^{-1}$, where $v_i^* \geq v_i$. For each credibility ceiling scenario, we estimate the summary OR according to random-effects calculations (i.e., the point estimate of the average effect). The estimated between-study heterogeneity would also have to change as it depends on the study-specific variances.

Choice of ceiling c values is arbitrary and may be based on knowledge of prior evidence or expert opinion. Different study designs and research questions should have different plausible ceilings. For this reason, we used here a range of plausible values so as to explore their impact on the results and inferences of the meta-analyses.

For each credibility ceiling scenario, we estimate the summary OR according to random-effects calculations, the extent of between-study heterogeneity (I^2 metric [22]), and the likelihood ratio (LR) of the OR exceeding vs. not exceeding 1.00. We also calculated the predictive intervals for each meta-analyses separately, which represent the range within which the effect estimate in a new study will lie [20]. The prediction interval for the effect size in a new study may be approximated by using the formula $\text{mean} \pm t_{\text{df}} \times \sqrt{\mathrm{var}(\widehat{\mu}) + \tau^2}$, where t is the appropriate centile point (e.g., 95%) of the t distribution with $k - 2$ degrees of freedom (k is the number of studies). Finally, we assessed whether the meta-analyses that retained nominal statistical significance ($P < 0.05$) despite increasing credibility ceilings were more likely to be those that had no hints of bias based on other tests that we had previously applied to these same meta-analyses [5]. These hints of potential bias included small-study effects [17], excess significance [18], and large between-study heterogeneity [19]. Small-study effects, meaning whether smaller studies give larger estimates of effect size compared with larger studies, were evaluated with the Egger regression asymmetry test [23]. Excess statistical significance was evaluated using the excess significance test [18,24], which evaluates whether the observed number of studies with nominally statistically significant results is too large based on the power that these studies have to detect plausible effects at $\alpha = 0.05$. Large between-study heterogeneity was defined as I^2 point estimates greater than 50% [19].

All calculations were performed in Stata 10 (StataCorp, 2007, Stata Statistical Software: Release 10, StataCorp LP, College Station, TX, USA).

3. Results

3.1. Description of the included meta-analyses

The selection of the included studies has been described in Tsilidis et al. [5]. In brief, 65 articles were reviewed in full text, and 37 articles were finally selected [25−61], which provided data on 98 meta-analyses (including 847 studies) in six broad fields of biomarkers and cancer risk [insulinlike growth factor (IGF)/insulin system ($n = 21$ comparisons), sex hormones ($n = 13$ comparisons), diet ($n = 31$ comparisons), inflammation ($n = 3$ comparisons), infectious agents ($n = 22$ comparisons), and environment ($n = 8$ comparisons)] (Table 1). The predictive intervals did not include the null value in 10 meta-analyses (two in IGF/insulin system, one in sex hormones, one in diet, and six on infectious agents).

3.2. Comparison of results with ceiling 0% and 10% for OR >1.00

As shown in Table 1, without applying correction for any ceiling, 43 meta-analyses (44%) had a statistically significant finding (34 showed increased risk of cancer). The proportion of statistically significant findings in each field was 86% on infectious agents, 67% on inflammation, 52% on IGF/insulin, 31% on sex hormones, 19% on diet, and 13% on environmental factors. The I^2 point estimate was >0 in 69 meta-analyses (70%).

As an illustration (Table 1), with an application of a typical ceiling of 10% for OR >1.00, only approximately half of the meta-analyses retained their statistical significance (21%), ranging from 64% on infectious agents to 0% on environmental factors and inflammation. The proportion of studies with I^2 point estimate >0 was also reduced to about half (36%).

3.3. Consideration of different ceilings

Table 2 summarizes the changes in the effect estimates, I^2, and LR for all scenarios of different credibility ceilings for all meta-analyses. Detailed information for changes in the summary OR, I^2, and LR for credibility ceilings 0.05−0.40 for OR >1 and for OR >1.2 is shown in Supplementary Figures 1−6, and Supplementary Table 1 at www.jclinepi.com summarizes data for each field (different types of risk factors).

Table 1. Analytical description of the 98 selected meta-analyses and predictive intervals with random effects (95% CI), I^2, and LR for credibility ceilings 0% and 10% under the scenario of the effect to be at the opposite direction

Meta-analysis (Ref)	Comparison	Studies	Cases/controls	Ceiling 0%			Ceiling 10%			Predictive intervals
				Random effects	I^2 (%)	LR	Random effects	I^2 (%)	LR	
IGF/insulin										
[32]	IGF-1 and CRC	11	2,862/4,966	1.07 (1.01, 1.14)	0	79	1.07 (1.00, 1.14)	0	43	1.00–1.16
[33]	IGFBP-3 and CRC	7	1,106/2,395	0.98 (0.64, 1.51)	60	1.15	0.97 (0.64, 1.45)	50	0.8	0.27–3.50
[33]	IGF-2 and CRC	3	384/1,301	1.95 (1.26, 3.00)	0	792	1.84 (1.03, 3.30)	0	48	0.12–32.2
[34]	IGF-1 and premenopausal BrCA	11	1,937/4,096	1.21 (0.98, 1.49)	29	26	1.12 (0.94, 1.33)	0	8	0.74–1.97
[34]	IGF-1 and postmenopausal BrCA	15	2,853/5,332	1.30 (1.13, 1.49)	0	> 1,000	1.20 (1.01, 1.40)	0	57	1.12–1.51
[34]	IGFBP-3 and premenopausal BrCA	11	1,915/4,012	0.99 (0.83, 1.19)	0	1.15	0.98 (0.81, 1.18)	0	0.7	0.80–1.22
[34]	IGFBP-3 and postmenopausal BrCA	15	2,816/5,196	1.22 (1.01, 1.49)	31	45	1.08 (0.91, 1.29)	0	4	0.75–2.00
[35]	IGF-1 and lung CA	6	1,043/11,472	0.98 (0.68, 1.41)	41	1.15	1.02 (0.71, 1.46)	41	1.1	0.38–2.53
[35]	IGFBP-3 and lung CA	6	1,043/11,472	0.96 (0.59, 1.56)	67	1.3	0.99 (0.60, 1.62)	57	0.9	0.21–4.35
[36]	IGF-1 and PrCA	42	7,481/11,866	1.21 (1.07, 1.36)	88	880	1.02 (0.95, 1.09)	46	2	0.58–2.51
[36]	IGF-2 and PrCA	10	923/1,874	1.17 (0.93, 1.47)	77	10	1.07 (0.87, 1.31)	47	3	0.55–2.50
[36]	IGFBP-3 and PrCA	29	6,676/10,484	0.88 (0.79, 0.98)	81	118	1.01 (0.95, 1.06)	24	1	0.52–1.47
[36]	IGF-1/BP-3 and PrCA	11	3,545/6,132	1.10 (0.97, 1.24)	80	15	1.01 (0.94, 1.09)	30	1.5	0.72–1.69
[36]	IGFBP-1 and PrCA	3	617/936	1.20 (0.65, 2.22)	92	2.5	0.84 (0.63, 1.12)	50	0.1	0.00–2,521
[36]	IGFBP-2 and PrCA	5	844/1,826	1.18 (0.90, 1.54)	78	7.5	0.97 (0.85, 1.10)	0	0.5	0.46–3.04
[37]	C-peptide and CRC	12	1,309/4,233	1.51 (1.14, 1.99)	54	501	1.20 (0.96, 1.51)	10	17	0.66–3.45
[37]	C-peptide and BrCA	11	1,403/2,114	1.35 (1.01, 1.81)	64	43	1.07 (0.90, 1.28)	0	3	0.53–3.41
[37]	C-peptide and endometrial CA	4	388/474	1.18 (0.57, 2.43)	69	2	0.89 (0.58, 1.35)	0	0.41	0.05–25.4
[37]	C-peptide and pancreas CA	2	209/483	1.70 (1.11, 2.61)	0	136	1.64 (0.95, 2.83)	0	24	Inestimable
[37]	Glucose and CRC	11	1,741/1.38 m	1.28 (1.06, 1.54)	47	190	1.12 (0.99, 1.27)	0	30	0.79–2.07
[37]	Glucose and pancreas CA	5	298/1.33 m	1.98 (1.67, 2.35)	0	> 1,000	1.80 (1.17, 2.77)	0	266	1.50–2.61
Sex hormones										
[38]	T and PrCA	17	3,886/6,438	0.98 (0.87, 1.10)	0	2	0.98 (0.86, 1.11)	0	0.5	0.86–1.11
[38]	Free T and PrCA	14	3,550/5,815	1.12 (0.98, 1.27)	0	21	1.10 (0.97, 1.26)	0	15	0.97–1.28
[38]	DHT and PrCA	7	1,010/1,445	0.88 (0.69, 1.11)	0	6	1.14 (0.90, 1.45)	0	6	0.64–1.20
[38]	A-diol G and PrCA	8	2,453/3,035	1.15 (0.95, 1.38)	24	12	1.08 (0.92, 1.28)	0	4.3	0.78–1.69
[38]	DHES-S and PrCA	7	1,154/1,870	1.29 (0.99, 1.68)	17	30	1.17 (0.92, 1.47)	0	9.4	0.76–2.17
[38]	D4 and PrCA	6	1,781/2,430	1.02 (0.85, 1.21)	0	1.3	1.02 (0.85, 1.21)	0	1.3	0.79–1.31
[38]	E2 and PrCA	9	2,186/3,039	0.92 (0.78, 1.09)	0	5.1	1.09 (0.92, 1.28)	0	5	0.75–1.12
[38]	Free E2 and PrCA	8	2,043/2,735	0.97 (0.82, 1.16)	0	1.6	1.00 (0.83, 1.19)	0	1	0.79–1.21
[38]	SHBG and PrCA	15	3,704/5,998	0.86 (0.76, 0.97)	0	149	1.14 (1.00, 1.30)	0	41	0.75–0.98
[39]	E2 and postmenopausal BrCA	9	656/1,709	1.26 (1.07, 1.49)	42	416	1.14 (0.97, 1.32)	0	17	0.83–1.92
[40]	2OHE1 and PrCA	2	122/414	0.76 (0.45, 1.28)	0	5.8	1.13 (0.78, 2.24)	0	6	Inestimable
[40]	16a-OHE1 and PrCA	2	122/414	1.82 (1.08, 3.05)	0	84	1.86 (0.94, 3.67)	0	26	Inestimable
[40]	2OHE1/16a-OHE1 and PrCA	2	122/414	0.52 (0.31, 0.89)	0	120	1.96 (0.94, 3.67)	0	27	Inestimable
Diet										
[41]	25(OH) vitamin D and colon CA	7	984/1,960	0.78 (0.53, 1.13)	46	9.6	1.14 (0.86, 1.52)	1	5	0.28–2.11
[42]	25(OH) vitamin D and PrCA	11	3,124/4,682	1.03 (0.95, 1.11)	26	3.1	1.02 (0.95, 1.09)	13	2	0.87–1.22
[43]	25(OH) vitamin D and BrCA	7	5,489/5,841	0.55 (0.38, 0.80)	86	> 1,000	1.23 (0.99, 1.53)	4	33	0.16–1.92
[43]	1a,25$(OH)_2$ vitamin D and BrCA	3	1,802/1,825	0.99 (0.68, 1.44)	47	1.1	1.00 (0.69, 1.44)	40	1	0.02–42.5
[44]	Carotenoids and lung CA	4	538/7,265	0.70 (0.44, 1.11)	45	15	1.26 (0.88, 1.79)	0	8	0.13–3.78
[44]	B-carotene and lung CA	10	2,846/34,783	0.84 (0.66, 1.07)	41	11	1.13 (0.87, 1.47)	31	5	0.46–1.54

(*Continued*)

Table 1. Continued

Meta-analysis (Ref)	Comparison	Studies	Cases/controls	Ceiling 0%			Ceiling 10%			Predictive intervals
				Random effects	I^2 (%)	LR	Random effects	I^2 (%)	LR	
[44]	A-carotene and lung CA	5	835/4,783	0.88 (0.59, 1.33)	53	2.6	1.00 (0.74, 1.33)	1	1	0.25–3.13
[44]	B-cryptoxanthin and lung CA	5	835/4,783	0.82 (0.40, 1.69)	75	2.3	1.16 (0.55, 2.42)	70	2	0.07–9.88
[44]	Lutein/zeaxanthin and lung CA	4	559/4,507	0.95 (0.67, 1.36)	11	1.5	1.00 (0.70, 1.42)	0	1	0.37–2.45
[44]	Lycopene and lung CA	4	727/4,567	0.71 (0.51, 0.99)	0	48	1.32 (0.91, 1.90)	0	13	0.34–1.46
[45]	Selenium and lung CA	6	1,029/1,658	0.77 (0.56, 1.08)	42	14	1.18 (0.86, 1.62)	20	6	0.33–1.84
[46]	SFA and BrCA	6	1,085/1,485	1.05 (0.79, 1.39)	0	1.7	1.02 (0.76, 1.36)	0	1	0.70–1.57
[46]	Stearic acid and BrCA	7	1,295/1,507	0.93 (0.69, 1.26)	14	2.1	1.03 (0.77, 1.37)	0	1	0.53–1.64
[46]	Palmitic acid and BrCA	7	1,295/1,507	1.05 (0.69, 1.58)	59	1.4	1.00 (0.67, 1.50)	51	1	0.31–3.52
[46]	MUFA and BrCA	5	931/1,360	1.44 (0.82, 2.53)	67	8.8	1.16 (0.79, 1.73)	22	3	0.22–9.54
[46]	Palmitoleic acid and BrCA	2	267/531	1.26 (0.41, 3.89)	81	1.9	1.12 (0.37, 3.35)	61	1	Inestimable
[46]	Oleic acid and BrCA	9	1,740/1,983	0.99 (0.70, 1.38)	70	1.14	0.92 (0.68, 1.24)	39	0.4	0.34–2.82
[46]	n-6 PUFA and BrCA	7	1,206/1,461	0.75 (0.53, 1.06)	16	18	1.24 (0.88, 1.75)	1	8	0.39–1.46
[46]	Linoleic acid and BrCA	8	1,449/1,632	0.85 (0.57, 1.26)	60	3.7	1.05 (0.73, 1.52)	42	1	0.27–2.73
[46]	Arachidonic acid and BrCA	5	975/1,251	0.89 (0.65, 1.22)	0	3.2	1.07 (0.77, 1.50)	0	2	0.54–1.48
[46]	n-3 PUFA and BrCA	8	1,360/1,586	0.79 (0.56, 1.11)	37	10	1.17 (0.85, 1.62)	19	5	0.34–1.84
[46]	α-Linolenic acid and BrCA	8	1,586/1,858	0.80 (0.59, 1.08)	39	12.3	1.11 (0.87, 1.42)	0	4	0.38–1.70
[46]	Eicosapentaenoic acid and BrCA	5	931/1,360	0.91 (0.87, 0.95)	0	>1,000	1.13 (0.99, 1.30)	0	29	0.84–0.97
[46]	Docosahexaenoic acid and BrCA	7	1,463/1,799	0.73 (0.53, 1.02)	36	30	1.17 (0.88, 1.55)	0	6	0.33–1.64
[47]	Enterolactone and BrCA	12	3,088/4,622	0.79 (0.61, 1.02)	71	29	1.00 (0.83, 1.20)	18	1	0.33–1.87
[48]	α-Linolenic acid and PrCA	6	1,091/1,270	1.54 (1.16, 2.06)	16	631	1.35 (1.01, 1.82)	0	45	0.87–2.75
[49]	Folate and BrCA	6	1,239/2,345	0.67 (0.46, 1.00)	41	41	1.30 (0.95, 1.79)	0	18	0.25–1.94
[50]	Folate and PrCA	7	2,958/6,962	1.11 (0.96, 1.28)	39	11	1.02 (0.96, 1.09)	0	3	0.79–1.55
[50]	Vitamin B12 and PrCA	6	2,906/6,495	1.10 (1.01, 1.19)	45	89	1.06 (1.00, 1.13)	0	39	0.89–1.36
[50]	Total homocysteine and PrCA	4	2,028/4,987	0.91 (0.70, 1.19)	14	3	1.05 (0.83, 1.33)	0	2	0.43–1.91
[51]	Vitamin B6 and CRC	4	883/1,424	0.52 (0.38, 0.71)	0	>1,000	1.85 (1.15, 2.98)	0	173	0.26–1.03
Inflammation										
[52]	C-reactive protein and CRC	8	1,159/37,986	1.12 (1.01, 1.25)	51	73	1.07 (0.98, 1.17)	22	13	0.85–1.49
[53]	C-reactive protein and CA	14	4,438/70,107	1.10 (1.02, 1.18)	73	227	1.03 (0.98, 1.08)	18	5	0.85–1.42
[53]	Interleukin-6 and CA	4	868/5,917	1.01 (0.90, 1.12)	21	1.2	1.01 (0.90, 1.11)	20	1	0.73–1.38
Infection										
[26]	*H pylori* and early gastric CA	15	2,722/13,976	3.38 (2.15, 5.32)	83	>1,000	1.71 (1.20, 2.46)	4	651	0.61–18.6
[54]	*H pylori* and gastric CA	15	2,284/2,770	2.29 (1.71, 3.05)	76	>1,000	1.45 (1.15, 1.82)	0	>1,000	0.77–6.82
[54]	cagA and gastric CA	10	1,707/2,124	2.87 (1.95, 4.22)	85	>1,000	1.74 (1.27, 2.38)	0	>1,000	0.73–11.3
[55]	*H pylori* and CRC	14	1,709/1,872	1.49 (1.16, 1.90)	58	>1,000	1.16 (0.97, 1.39)	0	18	0.67–3.30
[55]	*H pylori* and laryngeal CA	3	108/249	2.02 (1.27, 3.23)	0	646	1.87 (1.04, 3.35)	0	54	0.10–41.8
[56]	*H pylori* and esophageal adeno-CA	13	840/2,890	0.57 (0.47, 0.69)	15	>1,000	1.46 (1.15, 1.84)	0	>1,000	0.39–0.82
[56]	*H pylori* and ESCC	9	921/2,743	1.10 (0.78, 1.55)	73	2.4	1.02 (0.73, 1.43)	65	1	0.36–3.36
[56]	cagA and esophageal adeno-CA	5	275/1,197	0.41 (0.28, 0.62)	17	>1,000	1.84 (1.13, 3.02)	0	134	0.17–1.02
[56]	cagA and ESCC	4	483/1,844	1.01 (0.79, 1.27)	0	1.0	1.00 (0.79, 1.27)	0	1	0.60–1.69
[28]	*H pylori* and lung CA	4	199/231	3.24 (1.11, 9.41)	79	64	1.47 (0.86, 2.50)	0	12	0.03–371

(*Continued*)

Table 1. Continued

Meta-analysis (Ref)	Comparison	Studies	Cases/controls	Ceiling 0%			Ceiling 10%			Predictive intervals
				Random effects	I^2 (%)	LR	Random effects	I^2 (%)	LR	
[57]	HPV (DNA) and bladder CA	13	478/179	2.30 (1.33, 4.00)	6	663	1.81 (0.97, 3.37)	0	31	1.00–1.31
[57]	HPV (no DNA) and bladder CA	3	176/203	2.98 (1.65, 5.40)	0	>1,000	3.45 (1.06, 11.2)	0	50	0.06–140
[30]	HPV and oral CA	8	1,641/2,335	1.99 (1.17, 3.38)	62	178	1.41 (0.94, 2.13)	0	19	0.46–8.66
[30]	HPV and larynx CA	8	412/721	2.01 (0.96, 4.22)	50	30	1.49 (0.78, 2.82)	16	8	0.29–13.9
[30]	HPV and oropharynx CA	5	383/1,816	4.31 (2.07, 8.95)	56	>1,000	3.05 (1.27, 7.30)	0	161	0.49–38.2
[30]	HPV and tonsil CA	8	217/163	15.1 (6.78, 33.4)	0	>1,000	13.3 (3.04, 58.5)	0	>1,000	5.56–40.7
[31]	HPV and PrCA	9	2,242/2,622	1.52 (1.12, 2.06)	35	258	1.25 (1.00, 1.57)	0	39	0.74–3.09
[58]	HPV and cervical CA	12	589/3,068	8.08 (6.04, 10.8)	27	>1,000	6.93 (2.84, 16.9)	0	>1,000	4.22–15.5
[29]	HBV (HCV−) and liver CA	28	2,881/6,318	21.9 (14.9, 32.3)	86	>1,000	3.71 (2.16, 6.36)	0	>1,000	3.13–153
[29]	HCV (HBV−) and liver CA	26	2,079/5,615	20.3 (12.2, 33.7)	86	>1,000	7.04 (3.64, 13.6)	0	>1,000	1.68–245
[29]	HBV + HCV and liver CA	9	567/1,870	61.2 (27.0, 139)	37	>1,000	28.2 (4.79, 166)	0	>1,000	8.06–465
[61]	*T vaginalis* and cervical CA	2	642/65,122	1.88 (1.29, 2.74)	0	>1,000	1.70 (0.94, 3.06)	0	24	Inestimable
Environment										
[59]	Dieldrin and BrCA	5	1,550/1,673	1.15 (0.77, 1.69)	43	3	1.02 (0.73, 1.41)	0	1	0.37–3.52
[59]	Oxychlordane and BrCA	5	1,311/1,407	0.77 (0.51, 1.14)	51	9	1.18 (0.80, 1.73)	25	4	0.22–2.61
[59]	Cis-nonachlor and BrCA	3	692/695	1.09 (0.72, 1.64)	0	1.9	1.09 (0.72, 1.64)	0	2	0.08–15.7
[59]	Trans-nonachlor and BrCA	6	1,486/1,762	0.86 (0.68, 1.07)	0	10.2	1.14 (0.90, 1.45)	0	6	0.62–1.18
[60]	DDT and BrCA	24	5,222/6,147	0.97 (0.85, 1.11)	17	1.9	1.01 (0.89, 1.14)	5	1	0.72–1.32
[25]	DNA adducts and CA (current smokers)	8	509/407	3.76 (1.75, 8.05)	94	>1,000	1.44 (0.86, 2.40)	15	2	0.28–50.0
[25]	DNA adducts and CA (for smokers)	7	295/337	0.94 (0.71, 1.25)	0	2	1.07 (0.80, 1.41)	0	0.5	0.65–1.36
[25]	DNA adducts and CA (never smokers)	9	224/340	1.64 (0.72, 3.77)	79	7.3	0.91 (0.56, 1.46)	33	0.48	0.12–22.43

Abbreviations: CI, confidence interval; LR, likelihood ratio; IGF, insulinlike growth factor; CRC, colorectal cancer; IGFBP, insulinlike growth factor–binding protein; BrCA, breast cancer; CA, cancer; PrCA, prostate cancer; ESCC, esophageal squamous cell carcinoma; T, testosterone; DHT, dihydrotestosterone; A-diol G, androstanediol glucuronide; DHES-S, dehydroepiandrosterone sulfate; D4, androstenedione; E2, estradiol; SHBG, sex hormone–binding globulin; E1, estrone; SFA, total saturated fatty acids; MUFA, monounsaturated fatty acids; PUFA, polyunsaturated fatty acids; *H pylori*, *Helicobacter pylori*; HPV, human papillomavirus; HBV, hepatitis B virus; HCV, hepatitis C virus; *T vaginalis*, *Trichomonas vaginalis*; DDT, dichlorodiphenyltrichloroethane.

Random effects refers to summary odds ratio (95% CI) using the random-effects model.

With a ceiling of 20% for OR >1.0, only seven meta-analyses (7%) preserved statistical significance. Fifteen percent of the meta-analyses still had a nonzero point estimate of I^2. With a ceiling of 30% for OR >1.0, three meta-analyses (3%) retained statistical significance, and nine (9%) had a nonzero point estimate of I^2. For a ceiling of 40% for OR >1.0, six meta-analyses (6%) appear to have a statistically significant summary result. This spurious increase in the number of statistically significant meta-analyses occurred because with the reduction in the

Table 2. Changes in the proportion of studies with effect estimates >*X*, nominal statistical significance, and nonzero I^2 for all scenarios of the magnitude of the effect for all meta-analyses inclusively

All meta-analyses	Ceiling 0%	Ceiling 5%	Ceiling 10%	Ceiling 20%	Ceiling 30%	Ceiling 40%
X = 1, *n* (%)						
No. of meta-analyses with effect size >1.0	98 (100)	92 (94)	83 (85)	69 (70)	48 (49)	29 (30)
No. of meta-analyses with nominal statistical significance	43 (44)	34 (35)	21 (21)	7 (7)	3 (3)	6 (6)
No. of meta-analyses with nonzero I^2	69 (70)	57 (58)	35 (36)	15 (15)	9 (9)	1 (1)
X = 1.2, *n* (%)						
No. of meta-analyses with effect size > 1.2	54 (55)	50 (51)	42 (43)	33 (34)	25 (25)	20 (20)
No. of meta-analyses with nominal statistical significance	43 (44)	40 (41)	36 (37)	21 (21)	7 (7)	3 (3)
No. of meta-analyses with nonzero I^2	69 (70)	61 (62)	52 (53)	28 (29)	21 (21)	13 (13)

summary effect estimates, some of the meta-analyses gained statistical significance but on the opposite direction of the initial estimate; no meta-analysis retained statistical significance with an effect in the original direction with such a ceiling. Only one meta-analysis had a nonzero point estimate of I^2.

Considering various ceilings for OR >1.2, half of the meta-analyses (23%) lost statistical significance for a ceiling of 20%. Seven meta-analyses kept statistical significance for a ceiling of 30%, whereas only three meta-analyses survived the 40% ceiling. Between-study heterogeneity dropped gradually with higher ceilings (Table 2).

3.4. Meta-analyses with statistically significant results despite large ceilings

Table 3 lists the change in the summary results for different ceilings for each of the seven meta-analyses that had nominally statistically significant results when ceilings of 20% or higher are applied for OR >1.0, and Table 4 lists the changes in the effect estimates for each of the 21 meta-analyses that had nominally statistically significant results with the application of ceilings of 20% or higher for OR >1.2.

All seven meta-analyses that retained statistical significance with $c = 20\%$ for OR >1 were on infectious agents [cagA strains of *Helicobacter pylori* and gastric cancer, *H pylori* and esophageal adenocarcinoma, human papillomavirus (HPV) and tonsil/cervical cancer, and hepatitis B virus and/or hepatitis C virus (HCV) and liver cancer]. The meta-analysis on HPV and cervical cancer remained statistically significant even with 30% ceiling, whereas none survived a 40% ceiling.

The seven meta-analyses that retained statistical significance with $c = 30\%$ for OR >1.2 included three meta-analyses of infectious agents (cagA strains of *H pylori* and gastric cancer, *H pylori* and esophageal cancer, and HCV and liver cancer) but also four other meta-analyses [IGF-I and colorectal cancer, eicosapentaenoic acid and breast cancer, vitamin B12 and prostate cancer, and C-reactive protein (CRP) and cancer]. Two meta-analyses retained statistical significance with a ceiling of 40% (eicosapentaenoic acid and breast cancer and vitamin B12 and prostate cancer), whereas another one (IGFBP-3 and prostate cancer) gained statistical significance but with the effect being in the opposite direction.

3.5. Comparison against other bias-testing tools

Of the 24 statistically significant meta-analyses that had been previously identified [5] to have no signs of bias (no small-study effects, no large heterogeneity, and no excess of significant studies), the number that remained significant with $c = 5\%$, 10%, 20%, and 30% for OR >1.0 was 23 (96%), 15 (63%), 5 (21%), and 1 (4%), respectively. Of the 19 statistically significant meta-analyses that had been previously identified to have at least one such hint of potential bias, the respective numbers were 11 (58%), 6 (32%), 2 (11%), and 0 (0%). For OR >1.2, the number of meta-analyses that retained statistical significance for $c = 5\%$, 10%, 20%, 30% was 24 (100%), 22 (92%), 13 (54%), and 4 (17%) in the group without hints of bias and 16 (84%), 14 (74%), 8 (42%), and 3 (16%), respectively, in those with hints of bias. Thus, there was some suggestion that statistical significance was more likely to be retained with various ceilings in meta-analyses that had no hints of bias (e.g., $P = 0.048$ for $c = 10\%$ for OR >1.0). The difference was more prominent for ceilings related to OR >1.0 rather than OR >1.2.

3.6. Comparison against predictive intervals

Of the 10 meta-analyses with 95% predictive intervals excluding the null, the number that remained significant with $c = 5\%$, 10%, 20%, and 30% for OR >1.0 was 10 (100%), 9 (90%), 6 (60%), and 1 (1%), respectively. Of the 33 meta-analyses with 95% CIs excluding the null, but with 95% predictive intervals including the null, the respective numbers were 24 (73%), 12 (36%), 1 (3%), and 1 (3%). For OR >1.2, the number of meta-analyses that retained statistical significance for $c = 5\%$, 10%, 20%, and 30% was 10 (100%), 10 (100%), 10 (100%),

Table 3. Changes in summary effect estimates under the random-effects model for ceilings 0–40% for summary OR >1 for the seven studies that had a statistically significant summary effect for large ceilings

Comparison	Ceiling 0%	Ceiling 5%	Ceiling 10%	Ceiling 20%	Ceiling 30%	Ceiling 40%
cagA and gastric CA	2.87 (1.95–4.22)	1.74 (1.36–1.22)	1.74 (1.27–2.38)	1.74 (1.08–2.8)	NS	NS
H pylori and esophageal adeno-CA	1.77 (1.46–2.14)[a]	1.52 (1.23–1.89)	1.46 (1.15–1.84)	1.35 (1.02–1.79)	NS	NS
HPV and tonsil CA	15.1 (6.78–33.4)	14.46 (4.36–13.9)	13.34 (3.04–58.5)	11.6 (1.44–93.1)	NS	NS
HPV and cervical CA	8.08 (6.04–10.8)	6.93 (3.46–13.88)	6.93 (2.84–16.9)	6.93 (1.78–26.9)	6.96 (1.41–34.3)	NS
HBV (HCV−) and liver CA	21.9 (14.9–32.3)	9.38 (5.07–17.36)	3.71 (2.17–6.36)	2.69 (1.37–5.3)	NS	NS
HCV (HBV−) and liver CA	20.3 (12.2–33.7)	20.2 (12.2–33.7)	7.04 (3.64–13.6)	6.97 (2.58–18.8)	NS	NS
HBV + HCV and liver CA	61.2 (27.0–139)	61.2 (27.0–139.13)	28.2 (4.8–165.7)	28.2 (1.9–418)	NS	NS

Abbreviations: OR, odds ratio; CA, cancer; NS, not specified; *H pylori*, *Helicobacter pylori*; HPV, human papillomavirus; HBV, hepatitis B virus; HCV, hepatitis C virus.

[a] Effect inversed so that it is >1.00.

Table 4. Changes in summary effect estimates under the random-effects model for ceilings 0–40% for summary OR >1.2 for the 21 studies that had a statistically significant summary effect for large ceilings

Comparison	Ceiling 0%	Ceiling 5%	Ceiling 10%	Ceiling 20%	Ceiling 30%	Ceiling 40%
IGF-1 and CRC	1.07 (1.01–1.14)	1.07 (1.01–1.14)	1.07 (1.01–1.14)	1.07 (1.01–1.14)	1.07 (1.01–1.14)	NS
IGF-1 and postmenopausal BrCA	1.30 (1.13–1.49)	1.29 (1.12–1.48)	1.26 (1.09–1.46)	1.19 (1.02–1.40)	NS	NS
Glucose and CRC	1.28 (1.06–1.54)	1.21 (1.03–1.43)	1.15 (1.02–1.30)	1.12 (1.01–1.25)	NS	NS
Glucose and pancreas CA	1.98 (1.67–2.35)	1.81 (1.38–2.38)	1.78 (1.31–2.44)	1.77 (1.14–2.75)	NS	NS
SHBG and PrCA	1.17 (1.03–1.32)[a]	1.17 (1.03–1.32)	1.17 (1.03–1.32)	1.15 (1.01–1.30)	NS	NS
E2 and postmenopausal BrCA	1.26 (1.07–1.49)	1.26 (1.07–1.48)	1.22 (1.04–1.43)	1.17 (1.02–1.34)	NS	NS
25(OH) vitamin D and BrCA	1.82 (1.26–2.63)[a]	1.46 (1.11–1.91)	1.34 (1.05–1.70)	1.22 (1.01–1.47)	NS	NS
Eicosapentaenoic acid and BrCA	1.10 (1.06–1.15)[a]	1.10 (1.06–1.15)	1.10 (1.06–1.15)	1.10 (1.05–1.15)	1.10 (1.05–1.15)	1.10 (1.05–1.15)
Vitamin B12 and PrCA	1.10 (1.01–1.19)	1.09 (1.01–1.17)	1.07 (1.02–1.13)	1.07 (1.01–1.13)	1.06 (1.01–1.13)	1.06 (1.01–1.12)
Vitamin B6 and CRC	1.91 (1.40–2.61)[a]	1.91 (1.40–2.61)	1.87 (1.29–2.71)	1.82 (1.10–3.01)	NS	NS
C-reactive protein and CRC	1.12 (1.01–1.25)	1.12 (1.01–1.25)	1.12 (1.01–1.25)	1.12 (1.01–1.23)	NS	NS
C-reactive protein and CA	1.10 (1.02–1.18)	1.10 (1.02–1.18)	1.09 (1.02–1.17)	1.09 (1.02–1.16)	1.07 (1.01–1.14)	NS
H pylori and gastric CA	2.29 (1.71–3.05)	1.75 (1.42–2.17)	1.57 (1.29–1.90)	1.39 (1.10–1.76)	NS	NS
cagA and gastric CA	2.87 (1.95–4.22)	1.77 (1.44–2.19)	1.68 (1.34–2.09)	1.58 (1.21–2.07)	1.58 (1.03–2.43)	NS
H pylori and esophageal adeno-CA	1.77 (1.46–2.14)[a]	1.59 (1.30–1.95)	1.52 (1.22–1.88)	1.42 (1.12–1.79)	1.32 (1.01–1.72)	NS
cagA and esophageal adeno-CA	2.41 (1.61–3.61)[a]	2.02 (1.32–3.10)	1.92 (1.22–3.00)	1.75 (1.03–2.96)	NS	NS
HPV and tonsil CA	15.1 (6.78–33.4)	14.63 (4.71–45.4)	13.6 (3.36–55.1)	11.3 (1.65–77.2)	NS	NS
HPV and cervical CA	8.08 (6.04–10.8)	6.66 (3.58–2.4)	6.66 (3.0–14.78)	6.66 (1.98–22.42)	NS	NS
HBV (HCV−) and liver CA	21.9 (14.9–32.3)	10.6 (5.76–19.5)	4.0 (2.37–6.75)	2.55 (1.4–4.65)	NS	NS
HCV (HBV−) and liver CA	20.3 (12.2–33.7)	6.52 (4.08–10.4)	6.4 (3.55–11.4)	6.23 (2.61–14.9)	6.23 (1.54–25.1)	NS
HBV + HCV and liver CA	61.2 (27.0–139)	27.4 (7.5–100.6)	27.4 (5.2–145.4)	27.4 (2.16–347.6)	NS	NS

Abbreviations: OR, odds ratio; IGF, insulinlike growth factor; CRC, colorectal cancer; NS, not specified; BrCA, breast cancer; CA, cancer; SHBG, sex hormone–binding globulin; PrCA, prostate cancer; E2, estradiol; *H pylori*, *Helicobacter pylori*; HPV, human papillomavirus; HBV, hepatitis B virus; HCV, hepatitis C virus.

[a] Effect inversed so that it is >1.00.

and 3 (30%) in the group with 95% prediction intervals excluding the null and 30 (88%), 26 (79%), 11 (33%), and 4 (12%), respectively, in those with 95% prediction intervals including the null. Thus, meta-analyses with 95% predictive intervals excluding the null were more likely to retain nominal statistical significance with various ceilings (e.g., $P < 0.0001$ for $c = 20\%$ for OR >1.0).

3.7. A working example of applying the credibility ceiling method

To demonstrate the application of the method, we use the meta-analysis on IGF-2 and colorectal cancer [33] with synthesized ORs across three studies with a total sample size of 384 cases and 1,301 controls. The ORs (95% CI) in the three studies are 2.08 (0.85, 5.06), 2.09 (1.14, 3.82), and 1.59 (0.67, 3.76). The summary effect is statistically significant with an OR point estimate of 1.95 and 95% CI from 1.26 to 3.00 according to random effects model. We now reperform the meta-analysis by adjusting the variance of each study according to the credibility ceiling c using a range of different values for c. For example, for $c = 12\%$, the 95% CIs of the three studies become (0.61, 7.03), (0.61, 7.17), and (0.67, 3.76) and the summary OR and 95% CI become 1.82 (0.99, 3.35). For OR >1.2, the summary effect estimate statistical significance is lost for a larger ceiling of 19%

(1.82; 95% CI: 0.98, 3.35; Supplemental Figure 4 at www.jclinepi.com, graph number 2).

In the more general case, the code written in Stata to estimate the summary effect with different ceilings *c* is given by:

```
forvalues i=1/‘r(max)’ {
quietly metan logrevOR newse if compid==‘i’,nograph random eform
replace effectrandom2=r(ES) if compid==‘i’
replace seesrandom2=r(selogES) if compid==‘i’
replace effectrandom_l2=r(ci_low) if compid==‘i’
replace effectrandom_u2=r(ci_upp) if compid==‘i’
replace p_z2=r(p_z) if compid==‘i’
replace i_sq2=r(i_sq) if compid==‘i’
}
```

if prob=normal(z) and newvar=(logrevOR-logx)/zc1)^2 if prob < ceilingwhere prob = chance to observe a result on the opposite direction than the point estimate, logrevOR is the logarithm of the reversed OR and ceiling is the credibility ceiling we set for each study.

4. Discussion

Here, we have applied the credibility ceiling method to an array of meta-analyses on biomarkers and cancer risk to provide a critical approach to interpreting meta-analytic results in this field. Most of the meta-analyses that we examined lost their initial statistical significance for a credibility ceiling of 10% for OR >1.0. This ceiling is probably lenient, given the perplexity of the different types of bias an observational study may have. A value *c* = 10% merely states that we can reach even up to 90% certainty about the presence of a nonnull biomarker association just with a single study. Only seven biomarkers, all of them representing infectious agents maintained statistical significance for *c* = 20% for OR >1.0. This was supported also from the predictive intervals calculated for each meta-analysis. Most of the proposed cancer biomarkers of other noninfectious types may well be spurious.

Meta-analyses of IGF/insulin system lost most statistically significant associations with a ceiling of 10%. Meta-analyses examining the association between IGF-1 or IGFBP-3 and risk of prostate cancer [36] or C-peptide and breast cancer [37] lost their statistical significance even from a ceiling of 5%. One meta-analysis on IGF-I and postmenopausal breast cancer and one on glucose and pancreatic cancer had predictive intervals that did not include the null value. Some of these associations have been criticized for their magnitude as a survey has shown that authors of the primary studies on the association between IGF-I and prostate cancer supported the presence of strong evidence, whereas methodologists supported the presence of a null or a small effect [62].

Meta-analyses of inflammation biomarkers (CRP and colorectal [52] and other cancers [53]) also lost their statistical significance even with a ceiling of 5% for OR >1.0. This was reflected also in the predictive intervals. CRP appears to be an overrated biomarker for diverse outcomes, including both malignancies and cardiovascular outcomes, with no apparent causal relationship to any of them [6,7].

The fields of sex hormones, diet, and environment did not overall provide evidence of statistically significant associations. Only 1 of 13 and 1 of 31 had predictive intervals that did not include OR = 1. Practically, no study in the field of sex hormones and cancer risk survived a 10% ceiling. This is in agreement with a consortium-based analysis that also did not find any significant association between sex hormones and prostate cancer risk [38]. Nineteen percent of the meta-analyses on dietary associations had statistically significant findings, and of those, only two survived the 10% ceiling for OR >1.00 [48,51], both with a small number of synthesized studies and small number of cases. Large prospective cohort studies had little evidence to support these nutritional associations [63], and implausible results are common in published nutrition research [64,65]. Finally, in the area of environmental biomarkers, only one meta-analysis regarding DNA adducts and cancer risk in current smokers had a statistically significant association, but it did not survive a 5% ceiling for OR>1.0.

The field of infectious agents had robust evidence for the presence of not only significant associations with cancer but also large effects, which is one of the traditional criteria of causality [66,67]. However, several meta-analyses lost statistical significance with a ceiling of only 10% for OR >1.0 (e.g., *H pylori* and colorectal cancer, HPV and bladder/oral cancer, and *Trichomonas vaginalis* and cervical cancer). This may reflect the phenomenon where when a strong risk factor is documented (*H pylori* and stomach cancer and HPV and cervical cancer), other associations are proposed also on the same pathogen but different cancer type or other pathogens for the same cancer type. It is unclear whether the literature on these additional associations is reliable. For example, one may argue that some pathogens may have pleiotropic effects on multiple cancer sites. Conversely, bias may lead to the creation of a spurious literature where proposed associations “spill over” into other cancer types.

Credibility ceilings may be applied for different OR values. We observed that when ceilings were assumed for OR >1.2 rather than OR >1.0, the loss in statistically significant signals was less prominent. This is not surprising, because for the same *c* value, a ceiling for the presence of an effect OR >1.2 is a lesser constrain than the same ceiling for the presence of OR >1.0. With ceilings pertaining to OR >1.2, meta-analyses that include studies finding mostly small effects may not be as much affected and such associations may withhold even substantial ceilings, as shown by several examples of noninfectious biomarkers. However, it is unclear whether such small effects are genuine or easily distinguishable from noise. Biomarkers with OR <1.2 are unlikely to contribute much to risk of prognostic discrimination, but they can still create major debates about whether the estimated risks are real, let alone causal, even when large-scale evidence is accumulated [62].

In a previous evaluation, we had identified 24 meta-analyses that did not have hints of bias, including small-study effects, excess statistical significance, and large between-study heterogeneity [5]. These meta-analyses were more likely to retain statistical significance with the application of credibility ceilings than those that had at least one hint of bias. The difference was not very strong, however, suggesting that credibility ceilings offer independent information from what other tests of bias can offer. Moreover, we should acknowledge that no tests of bias can provide definite proof for (or exclude) bias and they may be sensitive to various assumptions [17,24,68,69]. Using the predictive interval approach, meta-analyses that did not include the null value in their predictive intervals were far more likely to retain statistical significance with increasing ceilings compared with those that included the null in their predictive interval.

Some additional limitations should be considered. First, the ceiling does not affect the point estimate of each study, so it is assumed that the point estimates are not biased. This is probably too conservative because in the presence of bias, it is expected that the point estimates would be inflated. However, the method aims simply at decreasing the credence that can be placed in any single study, not to correct it for bias. As mentioned earlier, there are other more complex methods [9−12,14,15] that change both the maximum likelihood estimate and standard error of the studies, but solid knowledge on what corrections to apply is often lacking. We have avoided the subjectivity of choosing only certain ceiling values, and instead, we applied a considerable range of ceilings and interpreted the results accordingly.

Second, the ceiling method inflates the variance to a greater extent in studies that have a point estimate away from the null. Studies with more extreme effects tend to obtain disproportionately greater variances, and as the ceilings become larger, the summary effects are more likely to get closer to the null.

Finally, the ceiling method will reduce the power of a meta-analysis to detect an effect, and this reduction can be considerable. However, it creates a framework where repeated replication of an association becomes more important. Nominal statistical significance of an association in ceiling-adjusted calculations requires the conduct of many studies showing a consistent effect. Especially in the field of biomarkers, their cost utility depends on their true discriminating ability, so biomarkers with the potential to be used in large populations should be carefully adopted based on robust evidence in multiple studies.

The application of this simple method with the aforementioned limitations may allow researchers to address in part the problems of bias in observational studies of biomarkers. Biomarkers are an appealing tool to use in clinical practice, and similar applications of the method in fields beyond cancer are warranted. Instead of trying to address the potential biases and confounding factors one by one (a formidable, if not impossible task), this method provides a skeptical approach to the observational associations of biomarkers in total and may help avoiding the propagation of unjustified claims in this literature.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2014.09.004.

References

[1] Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. Health Technol Assess 2003;7: iii-x, 1−173.

[2] Rifai N, Altman DG, Bossuyt PM. Reporting bias in diagnostic and prognostic studies: time for action. Clin Chem 2008;54:1101−3.

[3] Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ 2009;338:b606.

[4] Kyzas PA, Denaxa-Kyza D, Ioannidis JP. Almost all articles on cancer prognostic markers report statistically significant results. Eur J Cancer 2007;43:2559−79.

[5] Tsilidis KK, Papatheodorou SI, Evangelou E, Ioannidis JP. Evaluation of excess statistical significance in meta-analyses of 98 biomarker associations with cancer risk. J Natl Cancer Inst 2012; 104:1867−78.

[6] Ioannidis JP, Panagiotou OA. Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. JAMA 2011;305:2200−10.

[7] Ioannidis JP, Tzoulaki I. Minimal and null predictive effects for the most popular blood biomarkers of cardiovascular disease. Circ Res 2012;110:658−62.

[8] Ioannidis JP. Biomarker failures. Clin Chem 2013;59:202−4.

[9] Spiegelhalter DJ, Myles PJ. Evidence synthesis. In: Spiegelhalter DJ, Abrams K, Myles PJ, editors. Bayesian approaches to clinical trials and health-care evaluations. Chichester, New Hampshire, USA: Chichester John Wiley Sons Ltd; 2004.

[10] Lilford RJ, Braunholtz D. The statistical basis of public policy: a paradigm shift is overdue. BMJ 1996;313:603−7.

[11] Eddy DM, Hasselblad V, Shachter R. A Bayesian method for synthesizing evidence. The confidence profile method. Int J Technol Assess Health Care 1990;6:31−55.

[12] Wolpert RL, Mengersen KL. Adjusted likelihoods for synthesizing empirical evidence from studies that differ in quality and design: effects of environmental tobacco smoke. Stat Sci [Internet] 2004;19: 450−71.

[13] Greenland S. Multiple-bias modelling for analysis of observational data—discussion. J R Stat Soc Ser A Stat Soc [Internet] 2005; 168:267−306.

[14] Thompson S, Ekelund U, Jebb S, Lindroos AK, Mander A, Sharp S, et al. A proposed method of bias adjustment for meta-analyses of published observational studies. Int J Epidemiol 2011;40:765−77.

[15] Ioannidis JP. Commentary: adjusting for bias: a user's guide to performing plastic surgery on meta-analyses of observational studies. Int J Epidemiol 2011;40:777−9.

[16] Salanti G, Ioannidis JP. Synthesis of observational studies should consider credibility ceilings. J Clin Epidemiol 2009;62:115–22.
[17] Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011; 343:d4002.
[18] Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. Clin trials 2007;4:245–53.
[19] Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557–60.
[20] Higgins JP, Thompson SG. Presenting random effects meta-analyses: where are we going wrong? In 9th International Cochrane Colloquium. Lyon, France. 2001.
[21] DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin trials 1986;7:177–88.
[22] Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539–58.
[23] Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315: 629–34.
[24] Ioannidis JP. Clarifications on the application and interpretation of the test for excess significance and its extensions. J Math Psychol 2013;57:184–7.
[25] Veglia F, Loft S, Matullo G, Peluso M, Munnia A, Perera F, et al. DNA adducts and cancer risk in prospective studies: a pooled analysis and a meta-analysis. Carcinogenesis 2008;29:932–6.
[26] Wang C, Yuan Y, Hunt RH. The association between Helicobacter pylori infection and early gastric cancer: a meta-analysis. The Am J Gastroenterol 2007;102:1789–98.
[27] Zhuo XL, Wang Y, Zhuo WL, Zhang XY. Possible association of Helicobacter pylori infection with laryngeal cancer risk: an evidence-based meta-analysis. Arch Med Res 2008;39:625–8.
[28] Zhuo WL, Zhu B, Xiang ZL, Zhuo XL, Cai L, Chen ZT. Assessment of the relationship between Helicobacter pylori and lung cancer: a meta-analysis. Arch Med Res 2009;40:406–10.
[29] Donato F, Boffetta P, Puoti M. A meta-analysis of epidemiological studies on the combined effect of hepatitis B and C virus infections in causing hepatocellular carcinoma. Int J Cancer 1998;75:347–54.
[30] Hobbs CG, Sterne JA, Bailey M, Heyderman RS, Birchall MA, Thomas SJ. Human papillomavirus and head and neck cancer: a systematic review and meta-analysis. Clin Otolaryngol 2006;31: 259–66.
[31] Taylor ML, Mainous AG 3rd, Wells BJ. Prostate cancer and sexually transmitted diseases: a meta-analysis. Fam Med 2005;37:506–12.
[32] Rinaldi S, Cleveland R, Norat T, Biessy C, Rohrmann S, Linseisen J, et al. Serum levels of IGF-I, IGFBP-3 and colorectal cancer risk: results from the EPIC cohort, plus a meta-analysis of prospective studies. Int J Cancer 2010;126:1702–15.
[33] Morris JK, George LM, Wu T, Wald NJ. Insulin-like growth factors and cancer: no role in screening. Evidence from the BUPA study and meta-analysis of prospective epidemiological studies. Br J Cancer 2006;95:112–7.
[34] Key TJ, Appleby PN, Reeves GK, Roddam AW. Insulin-like growth factor 1 (IGF1), IGF binding protein 3 (IGFBP3), and breast cancer risk: pooled individual data analysis of 17 prospective studies. The Lancet Oncol 2010;11:530–42.
[35] Chen B, Liu S, Xu W, Wang X, Zhao W, Wu J. IGF-I and IGFBP-3 and the risk of lung cancer: a meta-analysis based on nested case-control studies. J Exp Clin Cancer Res 2009;28:89.
[36] Rowlands MA, Gunnell D, Harris R, Vatten LJ, Holly JM, Martin RM. Circulating insulin-like growth factor peptides and prostate cancer risk: a systematic review and meta-analysis. Int J Cancer 2009;124:2416–29.
[37] Pisani P. Hyper-insulinaemia and cancer, meta-analyses of epidemiological studies. Arch Physiol Biochem 2008;114:63–70.
[38] , Endogenous H, Prostate Cancer Collaborative G, Roddam AW, Allen NE, Appleby P, Key TJ. Endogenous sex hormones and prostate cancer: a collaborative analysis of 18 prospective studies. J Natl Cancer Inst 2008;100:170–83.
[39] Key T, Appleby P, Barnes I, Reeves G, Breast Cancer Collaborative G. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. J Natl Cancer Inst 2002;94:606–16.
[40] Barba M, Yang L, Schunemann HJ, Sperati F, Grioni S, Stranges S, et al. Urinary estrogen metabolites and prostate cancer: a case-control study and meta-analysis. J Exp Clin Cancer Res 2009;28:135.
[41] Yin L, Grandi N, Raum E, Haug U, Arndt V, Brenner H. Meta-analysis: longitudinal studies of serum vitamin D and colorectal cancer risk. Aliment Pharmacol Ther 2009;30(2):113–25.
[42] Yin L, Raum E, Haug U, Arndt V, Brenner H. Meta-analysis of longitudinal studies: serum vitamin D and prostate cancer risk. Cancer Epidemiol 2009;33:435–45.
[43] Chen P, Hu P, Xie D, Qin Y, Wang F, Wang H. Meta-analysis of vitamin D, calcium and the prevention of breast cancer. Breast Cancer Res Treat 2010;121(2):469–77.
[44] Gallicchio L, Boyd K, Matanoski G, Tao XG, Chen L, Lam TK, et al. Carotenoids and the risk of developing lung cancer: a systematic review. The Am J Clin Nutr 2008;88:372–83.
[45] Zhuo H, Smith AH, Steinmaus C. Selenium and lung cancer: a quantitative analysis of heterogeneity in the current epidemiological literature. Cancer Epidemiol biomarkers Prev 2004;13:771–8.
[46] Saadatian-Elahi M, Norat T, Goudable J, Riboli E. Biomarkers of dietary fatty acid intake and the risk of breast cancer: a meta-analysis. Int J Cancer 2004;111:584–91.
[47] Buck K, Zaineddin AK, Vrieling A, Linseisen J, Chang-Claude J. Meta-analyses of lignans and enterolignans in relation to breast cancer risk. The Am J Clin Nutr 2010;92:141–53.
[48] Simon JA, Chen YH, Bent S. The relation of alpha-linolenic acid to the risk of prostate cancer: a systematic review and meta-analysis. The Am J Clin Nutr 2009;89:1558S–64S.
[49] Larsson SC, Giovannucci E, Wolk A. Folate and risk of breast cancer: a meta-analysis. J Natl Cancer Inst 2007;99:64–76.
[50] Collin SM, Metcalfe C, Refsum H, Lewis SJ, Zuccolo L, Smith GD, et al. Circulating folate, vitamin B12, homocysteine, vitamin B12 transport proteins, and risk of prostate cancer: a case-control study, systematic review, and meta-analysis. Cancer Epidemiol biomarkers Prev 2010;19:1632–42.
[51] Larsson SC, Orsini N, Wolk A. Vitamin B6 and risk of colorectal cancer: a meta-analysis of prospective studies. JAMA 2010;303: 1077–83.
[52] Tsilidis KK, Branchini C, Guallar E, Helzlsouer KJ, Erlinger TP, Platz EA. C-reactive protein and colorectal cancer risk: a systematic review of prospective studies. Int J Cancer 2008;123:1133–40.
[53] Heikkila K, Harris R, Lowe G, Rumley A, Yarnell J, Gallacher J, et al. Associations of circulating C-reactive protein and interleukin-6 with cancer risk: findings from two prospective cohorts and a meta-analysis. Cancer Causes Control 2009;20:15–26.
[54] Huang JQ, Zheng GF, Sumanac K, Irvine EJ, Hunt RH. Meta-analysis of the relationship between cagA seropositivity and gastric cancer. Gastroenterology 2003;125:1636–44.
[55] Zhao YS, Wang F, Chang D, Han B, You DY. Meta-analysis of different test indicators: Helicobacter pylori infection and the risk of colorectal cancer. Int J colorectal Dis 2008;23:875–82.
[56] Islami F, Kamangar F. Helicobacter pylori and esophageal cancer risk: a meta-analysis. Cancer Prev Res 2008;1:329–38.
[57] Gutierrez J, Jimenez A, de Dios Luna J, Soto MJ, Sorlozano A. Meta-analysis of studies analyzing the relationship between bladder cancer and infection by human papillomavirus. The J Urol 2006; 176(6 Pt 1):2474–81; discussion 81.
[58] Mandelblatt JS, Kanetsky P, Eggert L, Gold K. Is HIV infection a cofactor for cervical squamous cell neoplasia? Cancer Epidemiol biomarkers Prev 1999;8:97–106.
[59] Khanjani N, Hoving JL, Forbes AB, Sim MR. Systematic review and meta-analysis of cyclodiene insecticides and breast cancer. J

Environ Sci Health C Environ Carcinog Ecotoxicol Rev 2007;25(1):23−52.

[60] Lopez-Cervantes M, Torres-Sanchez L, Tobias A, Lopez-Carrillo L. Dichlorodiphenyldichloroethane burden and breast cancer risk: a meta-analysis of the epidemiologic evidence. Environ Health Perspect 2004;112:207−14.

[61] Zhang ZF, Begg CB. Is Trichomonas vaginalis a cause of cervical neoplasia? Results from a combined analysis of 24 studies. Int J Epidemiol 1994;23:682−90.

[62] Panagiotou OA, Ioannidis JP. Primary study authors of significant studies are more likely to believe that a strong association exists in a heterogeneous meta-analysis compared with methodologists. J Clin Epidemiol 2012;65:740−7.

[63] Boyle P, Boffetta P, Autier P. Diet, nutrition and cancer: public, media and scientific confusion. Ann Oncol 2008;19(10):1665−7.

[64] Ioannidis JP. Implausible results in human nutrition research. BMJ 2013;347:f6698.

[65] Schoenfeld JD, Ioannidis JP. Is everything we eat associated with cancer? A systematic cookbook review. The Am J Clin Nutr 2013;97:127−34.

[66] Hill AB. The environment and disease: association or causation? Proc R Soc Med 1965;58:295−300.

[67] Guyatt GH, Oxman AD, Schunemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol 2011;64:380−2.

[68] Ioannidis JP, Patsopoulos NA, Evangelou E. Uncertainty in heterogeneity estimates in meta-analyses. BMJ 2007;335:914−6.

[69] Lundh A, Gotzsche PC. Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies. BMC Med Res Methodol 2008;8:22.